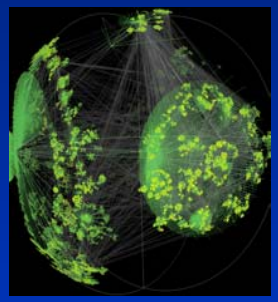
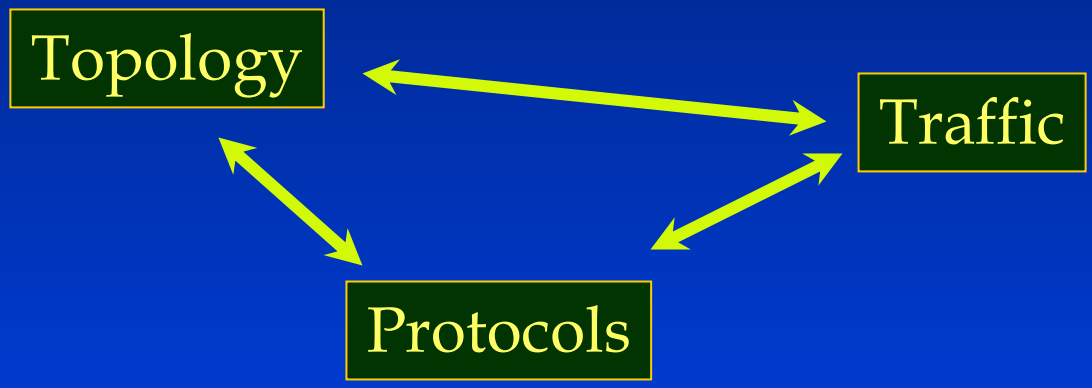
Analysis of the Internet Topology

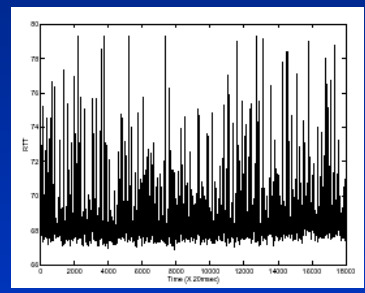
Michalis Faloutsos, U.C. Riverside (PI)
Christos Faloutsos, Carnegie Mellon U. (PI)
Thomas Karagiannis, (UCR)
Georgos Siganos, (UCR)

http://www.cs.ucr.edu/~michalis/PROJECTS/NMS/NMS.html

Big Picture: Simulate the Internet







- **Measure and model each component**
  - Identify simple properties and patterns
- **Model and simulate their interactions**

Overview of Research Directions

1. **Characterize and model network behavior:**
   - The use of Long Range Dependence
2. **Model and simulate BGP**
   - Large-Scale Realistic simulations (10,000 nodes)
3. **Model and simulate the Internet topology**
   - Identify topologies for simulations

Apply datamining techniques in network data

Progress Summary

- ✓ **Estimating Long Range Dependence**
  - There is no systematic way to estimate LRD
  - Pitfalls: estimators can be deceived
  - SELFYS software tool for performance analysis
- ✓ **A study of BGP policy routing robustness**
  - There is persistence and prevalence in BGP paths
  - Routing is fairly robust, but there is a lot of "noise" too
  - A data repository: 107Gb, 1 billion BGP paths
- **Measuring the performance of real-time applications**

Next Steps

- **A systematic approach to characterizing data:**
  - Can we characterize a trace with a few numbers?
  - Estimating LRD and quantifying burstiness
- **Effect of topology on performance**
  - What causes LRD in network behavior?
- **Realistic BGP simulation scenarios and models**
  - Simulation Models for BGP
  - Large Scale (10,000) BGP simulations
- **Spatiotemporal correlations: traffic and topology**

Work Integration and Collaboration

1. **Large Scale Measurements**
   - Nageswara Rao, ORNL
2. **Performance Characterization**
   - Rolf Riedi, Rice U.
3. **Large Scale BGP simulations**
   - George Riley Gtech, Dave Nichols, Dartmouth
   - Anja Feldmann, Saarland U.

Roadmap

I. **The hunt for Long Range Dependence**
   - How can I quantify LRD?

II. **BGP routing stability and policy**
   - How stable is BGP routing?

III. **Measurements for real-time applications**
   - How well does the network perform for i.e. VoIP?

Characterizing Network Behavior with Long Range Dependence

- LRD captures the "memory" of the behavior
- It is quantified by a single scalar number
- LRD appears in many aspects of networks
  - Traffic load, arrival times, delays, packet loss
- Open Question: what does it really tell us?

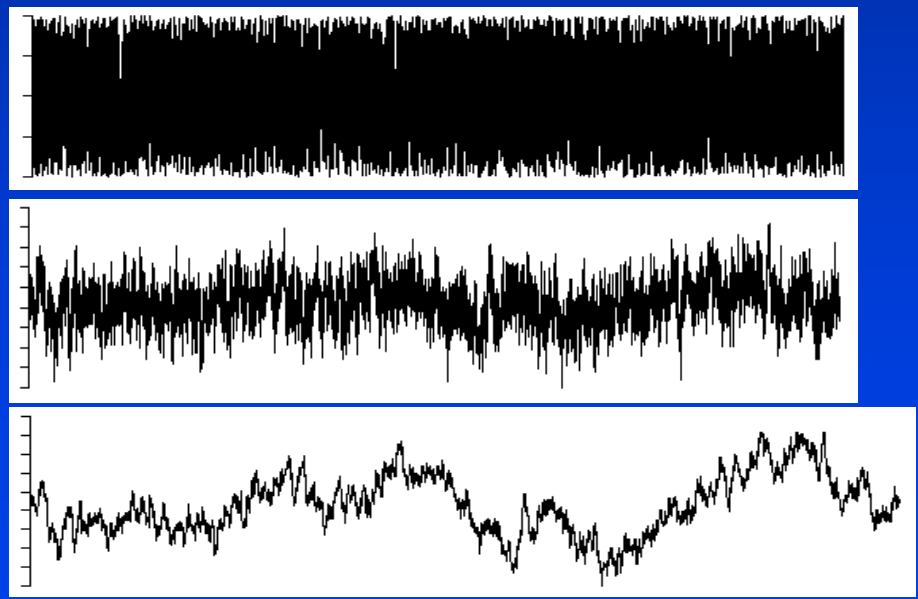
PROBLEM: We do not know how to estimate LRD!

- Many estimators and no systematic approach

The Intuition Behind LRD

- Capturing the "dependency" of the current measurement to previous values

- **White Noise**
- **Pink Noise**
- **Brownian Noise**

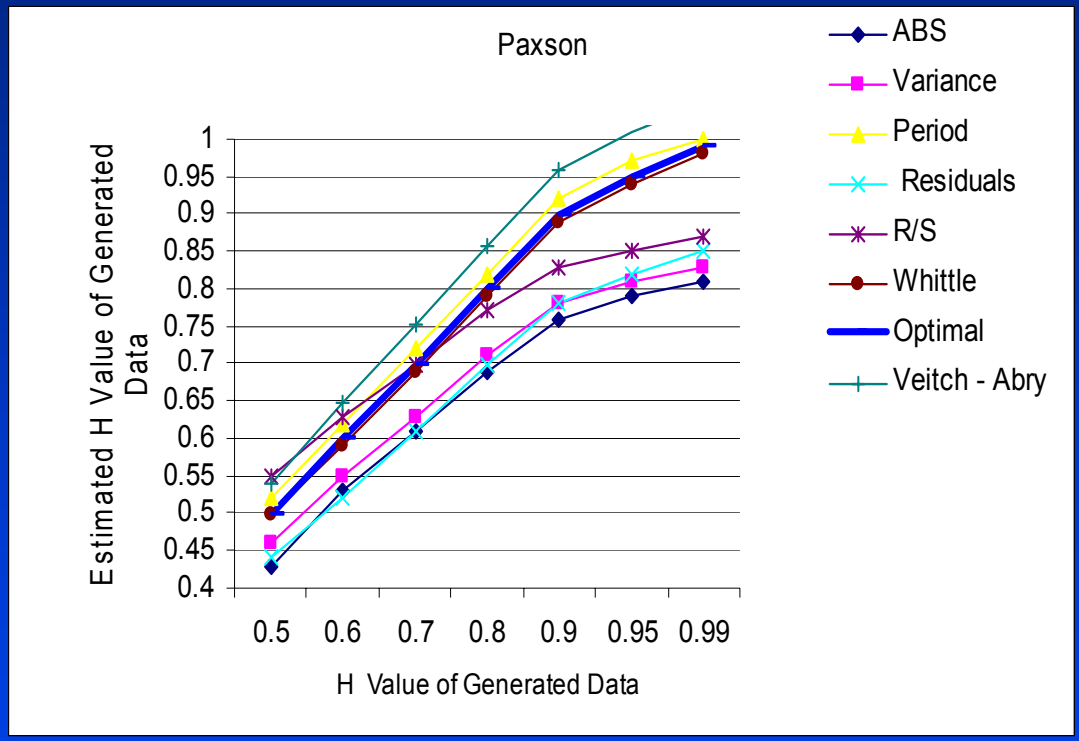


Idea: Reverse Engineering LRD

- Develop a library of behaviors of known data
- Three series of tests for the estimators

1. Evaluating the accuracy of the estimators
   - Synthetic Fractional Gaussian Noise (FGN)
2. Deceiving the estimators with non-LRD data
   - Periodicity, Noise, Trend
3. Applying the estimators on real data
   - Characterizing delay and packet loss

1. Accuracy: Synthetic LRD Data



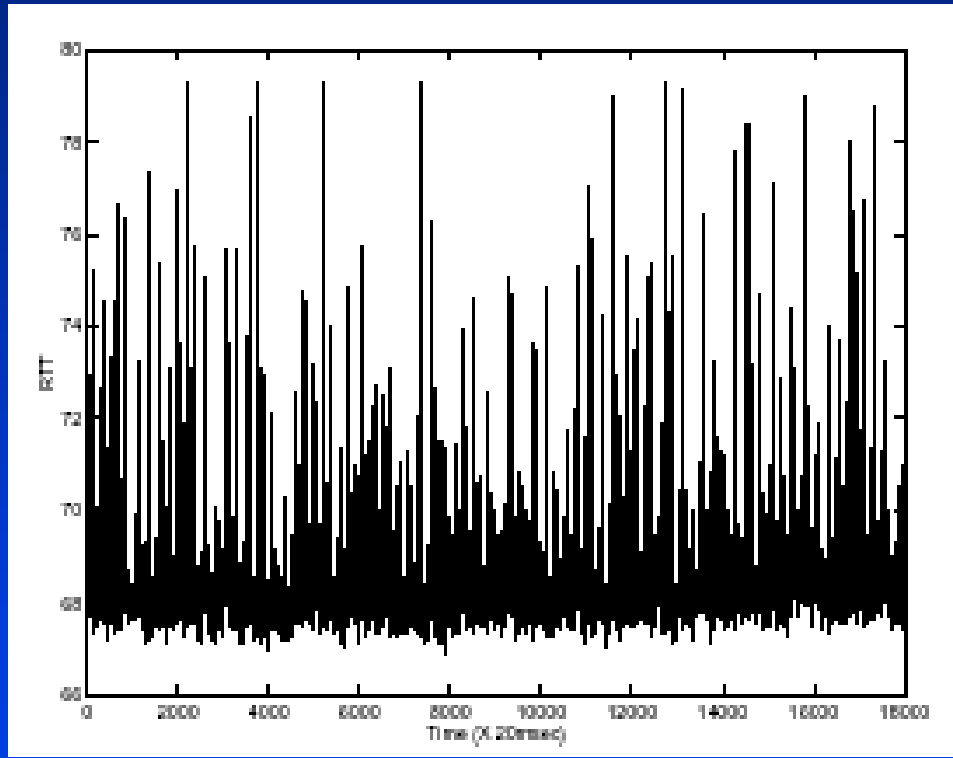
- Large difference in values!
- The Whittle and Periodogram are most accurate
- The rest can be significantly inaccurate!

2. Robustness: Deceiving the Estimators

- **Periodicity fools many estimators**
  - The Whittle, the Periodogram, the R/S and the Abry-Veitch falsely report LRD in series constructed by cosine functions and noise.
- **Noise affects the accuracy of estimation**
- **Trend (Non-Stationarity) also**
  - Whittle and Periodogram falsely report LRD

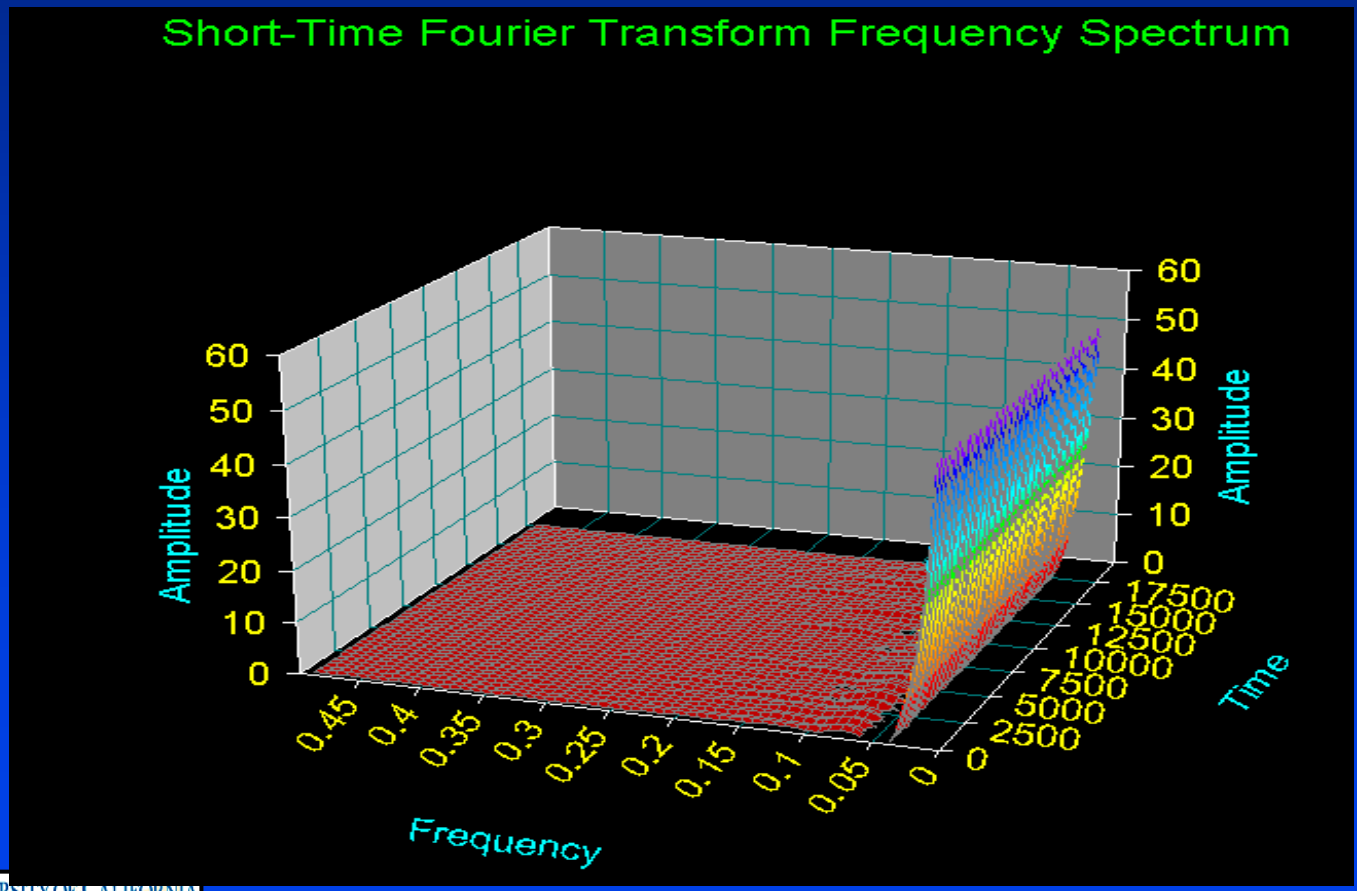
3. Analyzing Real Data

- Measured round trip time
- Initial signal does not exhibit LRD
- What do we do next?



The Measured Data Is Periodic

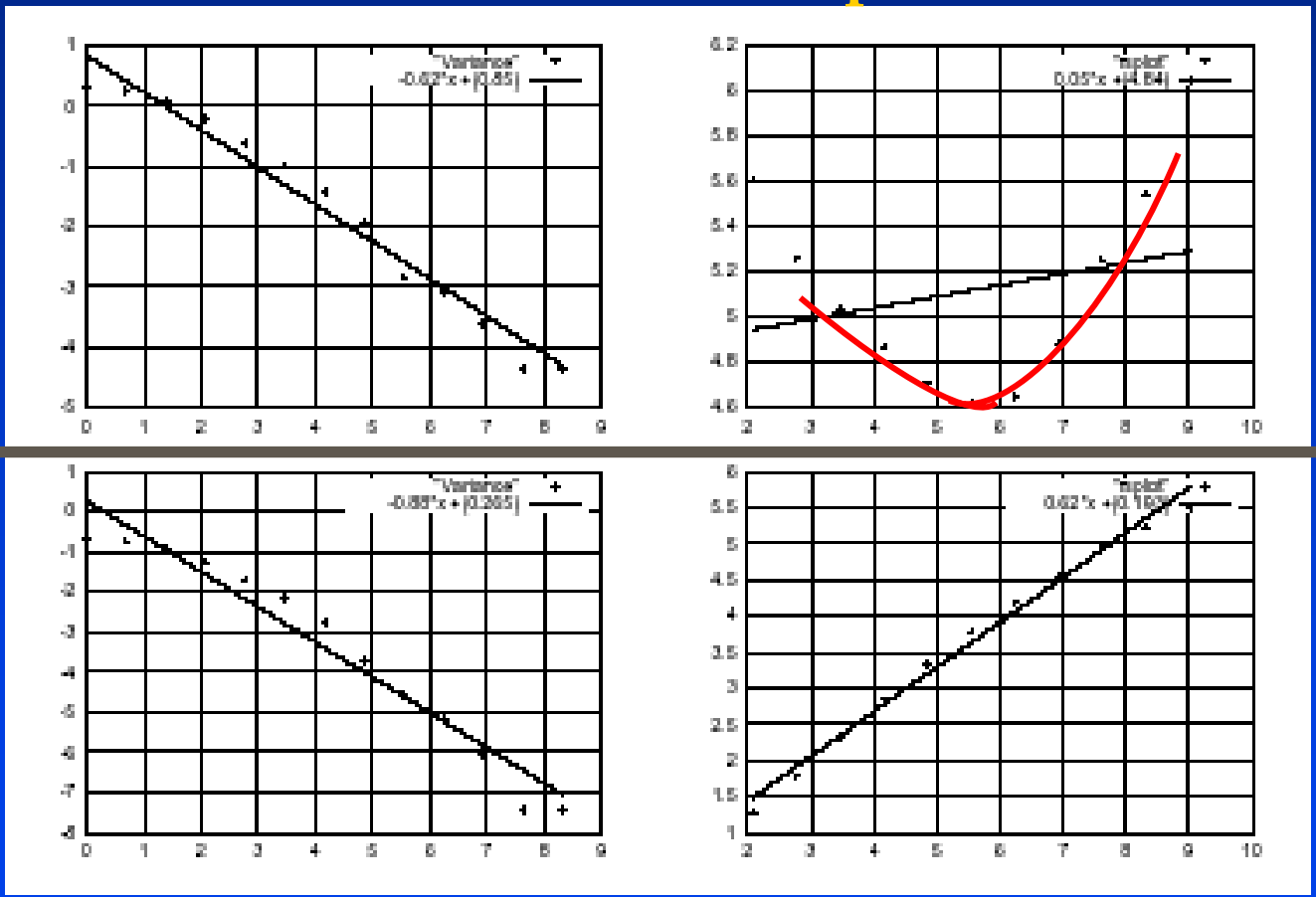
- There is periodicity throughout the dataset



Periodicity Hides the LRD!

Variance Method **RS-plot Method**

Measured (periodicity)



Without periodicity

Estimated: 0.55 and 0.68

Practical Lessons

- **LRD estimation and method must be reported**
- **LRD may exist even if all estimators do not agree**
- **There is no "consistent-winner" estimator**
  - We need to consult many of them
  - Drawing a conclusion is not obvious
- **Estimation can be thrown off by**
  - Noise, trend and periodicity

Towards a Systematic Approach

- Decompose signal and characterize each component separately
- Use all estimators
- Try reverse engineering

The SELFYS Tool

- **Given a trace**
  - Cleans data
  - Wavelet and Fourier analysis
  - Runs all LRD estimators
  - Plots results
- **SELFYS an open software reference point:**
  - Java
  - Modular
  - Free

http://www.cs.ucr.edu/~michalis/PROJECTS/NMS/NMS.html

Roadmap

I. **The hunt for Long Range Dependence**
   - How can I calculate LRD?

II. **BGP routing stability and policy**
   - How stable is BGP routing?

III. **Measurements for real-time applications**
   - How well does the network perform for VoIP?

BGP Routing/Policy Analysis

- **Overarching Goal:**
  - Develop a realistic detailed model for large scale realistic simulations
- **Now: A study of BGP routing robustness**
  - Persistence and prevalence of paths
  - Stability of advertisements
- **Next step:**
  - Study the customer-provider relationships

Using Massive BGP Routing Data

- **We use data from NLANR for almost 3 years**
  - Late 1997 to early 2001
- **Daily snapshots of BGP routing tables**
- **Created a database to facilitate path queries**
  - 107Gb of data, 1 billion BGP paths

Overview of Results for BGP Routing

Stable and persistent routing with some "noise"

- 44% prefixes are advertised for < 30 days
- 50% prefixes have a dominant path 84% of time
- 35% of prefixes use one path continuously for 90% of their time!
- Significant path multiplicity due to traffic engin.

Roadmap

I. **The hunt for Long Range Dependence**
   - How can I calculate LRD?

II. **BGP routing stability and policy**
   - How stable is BGP routing?

III. **Measurements for real-time applications**
   - How well does the network perform for VoIP?

Measurements: The Death of the Symmetry Assumption

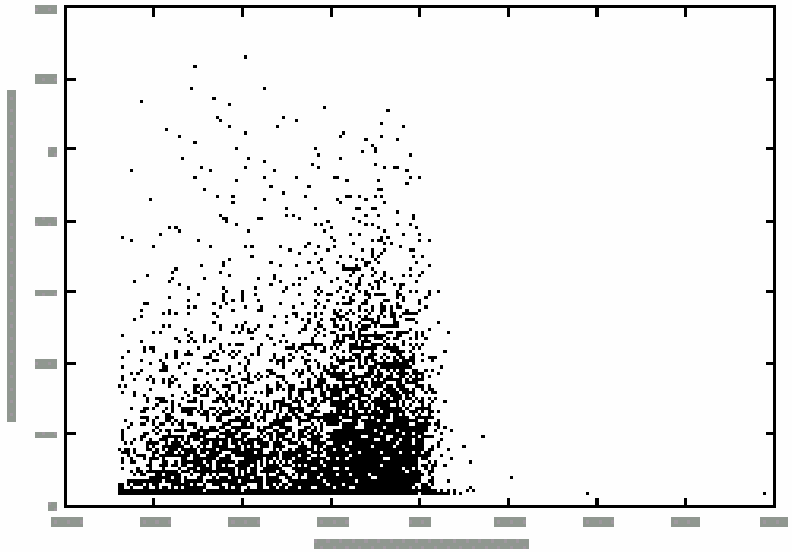


Fig. 1. CMU dataset with 40-msec sending rate and 320-byte packet size.

- One-way delay:

  Forward can be 10 times higher than backward delay

Overview of Contributions

- **Estimating Long Range Dependence**
  - There is no systematic way to estimate LRD
  - Pitfalls: estimators can be deceived
  - SELFYS software tool for performance analysis
- **A study of BGP policy routing robustness**
  - There is persistence and prevalence in BGP paths
  - Routing is fairly robust, but there is a lot of "noise" too
  - A data repository: 107Gb, 1 billion BGP paths
- **Measuring the performance of real-time applications**

http://www.cs.ucr.edu/~michalis/PROJECTS/NMS/NMS.html

Analysis of the Internet Topology

- Michalis Faloutsos, U.C. Riverside (PI)
- Christos Faloutsos, Carnegie Mellon U. (PI)

http://www.cs.ucr.edu/~michalis/PROJECTS/NMS/NMS.html